Pergamon

PII: S0042-6989(97)00008-4

Vision Res., Vol. 37, No. 23, pp. 3385–3398, 1997

Printed in Great Britain
0042-6989/97 $17.00 + 0.00

Origins of Scaling in Natural Images

DANIEL L. RUDERMAN*†

Received 16 July 1996; in revised form 10 November 1996

One of the most robust qualities of our visual world is the scale invariance of natural images. Not only has scaling been found in different visual environments, but the phenomenon also appears to be calibration-independent. This paper proposes a simple property of natural images which explains this robustness: they are collages of regions corresponding to statistically independent "objects". Evidence is provided for these objects having a power-law distribution of sizes within images, from which follows scaling in natural images. It is commonly suggested that scaling instead results from edges, each with power spectrum $1/k2$. This hypothesis is refuted by example. © 1997 Elsevier Science Ltd.

Natural images Power spectrum Scaling

INTRODUCTION

The properties of our visual world greatly influence the design of creatures' visual systems. Most fundamental, perhaps, is the amount of light available for vision. Vertebrate eyes which are adapted to dim environments tends to be larger, have a greater predominance of rods over cones, and achieve greater numerical apertures than those adapted to photon-rich settings.‡ However, the most important aspect of these luminous environments is the pattern of light fluctuations, both in space and time, which reach the eye. It is in this signal that information about the environment is conveyed. The eye must not only be able to count photons; it must also reliably detect variations in this count. And, just as the number of photons available to the eye determines its design, so do we expect that the statistical properties of these images select basic aspects of the pattern processing performed by the visual system.

This viewpoint has led to two interrelated areas of study in recent years. The first is in understanding how the statistical properties of natural images enter into the optimization of a visual system. By now, a large list of contributions has appeared using the specific optimization criteria of redundancy minimization or decorrelation (Laughlin, 1981; Srinivasan *et al.*, 1982; Atick & Redlich, 1990 Atick & Redlich, 1992; Atick, 1992; Atick *et al.*, 1992; Li & Atick, 1994; Dong & Atick, 1995b), maximization of information transmission (Bialek *et al.*, 1991; van Hateren, 1992a,b; Laughlin, 1992), sparseness of the neural encoding (Field, 1994; Olshausen & Field, 1996), and minimizing reconstruction error (Linsker, 1994; Ruderman, 1994a). This body of work has demonstrated that simple optimization principles combined with knowledge of image statistics can predict visual processing strategies which are found in nature.

The second research area is in the statistical characterization of the natural images themselves. Of great interest has been the discovery of scaling in these pictures (Burton & Moorhead, 1987; Field, 1987; Tolhurst *et al.*, 1992; van Hateren, 1992a; Ruderman & Bialek, 1994a,b). The key result states that the power spectrum of natural scenes§ takes the form of a power-law in the spatial frequency:

$$S(k) = A/k^{2-\eta}, \qquad (1)$$

where k is the magnitude of the spatial frequency, η is the "anomalous exponent" (usually small), and A is a constant which determines the overall image contrast. Our results (Ruderman & Bialek, 1994b) have revealed scaling over nearly three decades in spatial frequency we measured (up to approximately 30 cycles/deg), with $\eta = 0.19$. Interestingly, Dong & Atick (1995a) have shown this scaling to continue also in the time domain, which will be discussed later. Scaling in natural images also presents itself in statistics beyond the second-order measure given by the power spectrum; the shapes of entire histograms of pixel values and other statistics are invariant to changes in scale (Ruderman & Bialek, 1994b; Ruderman, 1994b), which demonstrates scaling to higher than just second-order.

That the process of geological formation of hillsides and valleys, or the structure of forests due to the succession of flora, can exhibit scaling through their images is perhaps not altogether surprising. Many natural

*The Physiological Laboratory, Downing Street, Cambridge CB2 3EG, U.K.

†Address all correspondence to present address: The Salk Institute, 10010 North Torrey Pines Road, La Jolla, CA 92037-1099, U.S.A. [*Email*: ruderman@salk.edu].

‡For a thorough and fascinating review of visual adaptations, see the work of Lythgoe (1979).

§This relation is only valid for the orientationally averaged spectrum. The spectrum itself has a non-trivial orientation dependence (Field, 1987; Ruderman, 1994b; van der Schaaf & van Hateren, 1996).

processes such as diffusion-limited growth formation (Vicsek, 1992), hard turbulence (Procaccia, 1984; Zocchi *et al.*, 1990), earthquakes (Turcotte, 1995), and even the large-scale structure of the universe (Coleman & Pietronero, 1992) produce scale-invariant patterns, with the universal presence of power-law correlations. It is striking, however, that the natural image datasets in which scaling was found are all quite different. No two sets of pictures were even from the same environment. Also, the methods of image capture differed not only in the choice of pictures taken, but in the spectral sensitivity of the medium (e.g., film or charge coupled device detectors) as well. Finally, the image data vary from raw, uncalibrated pixel values, to calibrated luminance, to the logarithm of the luminance. Despite these vast experimental differences in data collection the result is the same: natural images have scaling power spectra.

The purpose of this paper is to explain the ubiquity of scaling in natural image data. The underlying reason must be robust to calibration changes and it must manifest itself in widely varying image environments. To devise an answer, the basic structure of images found in natural environments was appealed to. Since the power spectrum is a statistical quantity, its behavior will follow from the statistics of the contributing images. Heuristically, we can liken these images to collages of patches corresponding to different "objects", which can appear in many combinations under varying lighting conditions (Ruderman, 1989, 1996; Baddeley, 1994). In the view of Helmholtz, objects contribute to the "ordinary conditions of vision", which occur when

> ... the visual organ is stimulated by light from outside; this outside light, coming from the opaque objects in its path that were the last to be encountered, and having reached the eye along rectilinear paths through an uninterrupted layer of air (von Helmholtz, 1881).

For the purposes of this paper the notion of an object is defined in a statistical sense—when items tend to co-appear—rather than a semantic one.

A given object contributes a finite set of pixels to an image. We expect that those pixels which belong to different objects will have less statistical dependence than pixels from a single object. Ultimately these statements can be quantified, as they will be shortly. The notion of statistically correlated and uncorrelated regions within images corresponding to objects provides a simple, robust path to scale invariance, as long as those objects appear in all sizes according to a power-law distribution.

The paper is organized as follows. In *"Scaling in natural images: the evidence"* the now classic result of scaling in the power spectrum is presented. To form a more intuitive approach the spectral scaling is Fourier transformed back to the spatial domain, where an equivalent result for the correlation function is derived. An explanation for the robustness of scaling in the correlation function (and thus the spectrum) is proposed in the section entitled *"Underlying regularity: objects and edges"*. It appeals to the notion of objects and statistical variations between them. Data are presented which support the notion of vastly different statistics between pixels from different objects than within a single object. Calibration independence of image scaling is demonstrated, supporting the hypothesis that the presence of overlapping objects is the dominant contribution to the second-order statistics. A simple two-dimensional model is developed in the section entitled *"Static images"* which captures all the essential ingredients of the scaling model, most notably occluding objects. Finally, the section entitled *"Moving images"* extends the results to the temporal domain.

SCALING IN NATURAL IMAGES: THE EVIDENCE

Let me begin by showing an example image from our database. The photograph in Fig. 1 was gathered in a central New Jersey forest during the spring of 1992 (for details of the imaging procedure see Ruderman & Bialek, 1994b). The image dataset used in this paper consists of 45 pictures of similar subject matter, consisting mainly of trees, rocks, and a stream. The images are 256 pixels on a side, and measure about 15 deg in visual angle. The natural logarithm of the calibrated luminance is used as image data, minus a constant for each image which gives the data a mean of exactly zero for each image. This is similar to an image-wide luminance normalization procedure.

The power spectrum

Our data (Ruderman & Bialek, 1994b) show scaling in the power spectrum of the form:

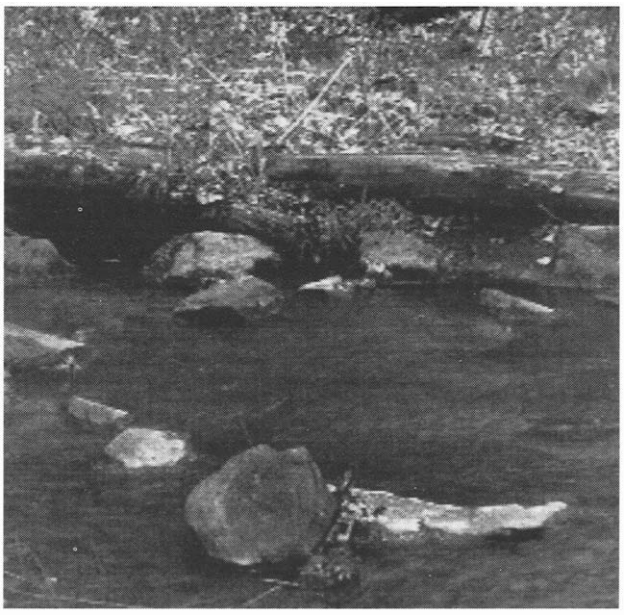

FIGURE 1. Example image from the dataset. All images were 256×256 pixels, measuring 15° of visual angle on a side. Data were taken as the logarithm of each pixel's luminance.

$$S(k) = A/k^{2-\eta}, \quad (2)$$

where k is the spatial frequency (measured, for example, in cycles/deg), A is a constant representing the overall contrast power in the images, and for our data $\eta = 0.19$ giving an overall exponent of 1.81. Images of waves and shoreline from the beach give an $\eta \approx -0.3$ (Ruderman & Bialek, unpublished). These data compare well with the work of others (Carlson, 1978; Burton & Moorhead, 1987; Field, 1987; Tolhurst *et al.*, 1992; Field, 1993; van der Schaaf & van Hateren, 1996), all of whom find a power-law of some exponent close to 2, though each uses a different set of data and varying measures of calibration.*

The reasons for measuring all statistics of the power spectrum are two-fold. Power spectra are a function of spatial frequency. Since the mid-1960s much of the characterization of visual systems has been done with respect to sinusoidal stimuli of a given spatial frequency (Braddick *et al.*, 1978). The contemporary visual scientist's understanding of vision is grounded largely in ideas about responses and sensitivities as a function of spatial frequency. Discussing natural stimuli in terms of spatial frequency content follows directly. Spectra also take a fundamental position in the statistical line-up since they embody all the second-order correlations in a translation-invariant ensemble of signals, which natural images are widely presumed to be.† Many of the optimization problems presented in the Introduction characterize natural images solely by their power spectrum. It has thus taken a prominent role in theories of sensory encoding. A notable exception is the work of Field (1994) and Olshausen & Field (1996), which relies on higher-order statistics.

Scaling in space

The Fourier analysis of images associated with measuring power spectra is not necessarily conducive to understanding which properties of natural images make them scale. Our grasp of image structure is best held in the spatial domain. Objects, after all, are generally spatially cohesive. In the Fourier domain, though, they spread and superpose over many frequency bands. Furthermore, the all-important process of occlusion, which occurs in projections of the three-dimensional world onto two-dimensional images, is not easily conceived of in relation to spatial frequency components. For these reasons I prefer, for the remainder of this paper, to work in the spatial domain. The advantages will become especially apparent when working with model image ensembles.

To start, the scaling result of Eq. (1) can be expressed equally well in the spatial domain in terms of the usual correlation function $C(x)$ as:

$$C(x) = -C_1 + C_2 x^{-\eta}, \quad (3)$$

where x is the separation distance between two pixels in an image, η assumes the same value as it did in the power spectrum, and C_1 and C_2 are positive (negative) constants for η positive (negative).‡ This expression is derived in the Appendix from the simple Fourier relation between the two-point correlation function and the power spectrum.

This correlation function is found experimentally as the expected product of the image at two pixels separated by a distance x:

$$C(x) = \langle\langle\langle\phi(\mathbf{x}_0)\phi(\mathbf{x}_0 + \mathbf{x})\rangle_\theta\rangle_{\mathbf{x}_0}\rangle_\phi. \quad (4)$$

Here $\phi(\mathbf{x})$ is the image value at position $\mathbf{x}$ (a two-vector), and the triple expectation value is over (from the outside inwards) all images ϕ, all initial positions $\mathbf{x}_0$, and all displacement vectors $\mathbf{x}$ of length x (parameterized by the angle θ). This this large expression can be rewritten schematically as $C(x) = <\phi(0)\phi(x)>$, with the ensemble average, shift over positions, and average over angles implied.

In preference to the correlation function another quantity, the "difference function", is used:

$$D(x) = \langle|\phi(0) - \phi(x)|^2\rangle, \quad (5)$$

with the meaning of the expectation value as already stated. The advantage of this formulation is that any unknown additive image offset does not enter the statistic. As mentioned above, a constant was added to all log-intensity image values in the database in order to make the average image value zero. This arbitrary constant does not appear in $D(x)$ since it only measures pixel differences. Equivalently, this statistic tabulates the expected values of the squared logarithm of intensity ratios within images. Using intensity ratios prevents the overall image luminance from entering the image data. By expanding the square in Eq. (5) and taking the expectation value, it can easily be seen that $D(x)$ is linearly related to $C(x)$ and takes the form:

$$D(x) = D_1 - D_2 x^{-\eta}, \quad (6)$$

with D_1 and D_2 positive (negative) for η positive (negative).

A power-law spectrum thus yields a power-law difference function, except for the presence of an added constant. $D(x)$ was measured by selecting 10^5 point pairs at random from each image and tabulating their joint statistics. In the resulting data the best choice for the constant D_1 was found so that the resulting difference gave the best power-law (of any exponent) by minimizing the mean-squared error between a power-law fit and the "residual function" $R(x) = D_1 - D(x)$. The best fit was found for $D_1 = 0.79$ and $D_2 = 0.64$; the resulting power-law residual had a slope of $\eta = 0.19$, matching the anomalous dimension from the power spectrum. This residual function is plotted in Fig. 2, showing a

*None of the authors has reported the details of the spectral response of their image detector, and so the precise nature of the measurements is unknown for the purposes of comparison.

†Since sky occurs more often at the top of image than at the bottom, this assumption can be brought into question. The answer to how good an approximation stationarity is must be answered quantitatively.

‡The case of $\eta = 0$ gives a logarithm instead of a power-law in x; see the Appendix.

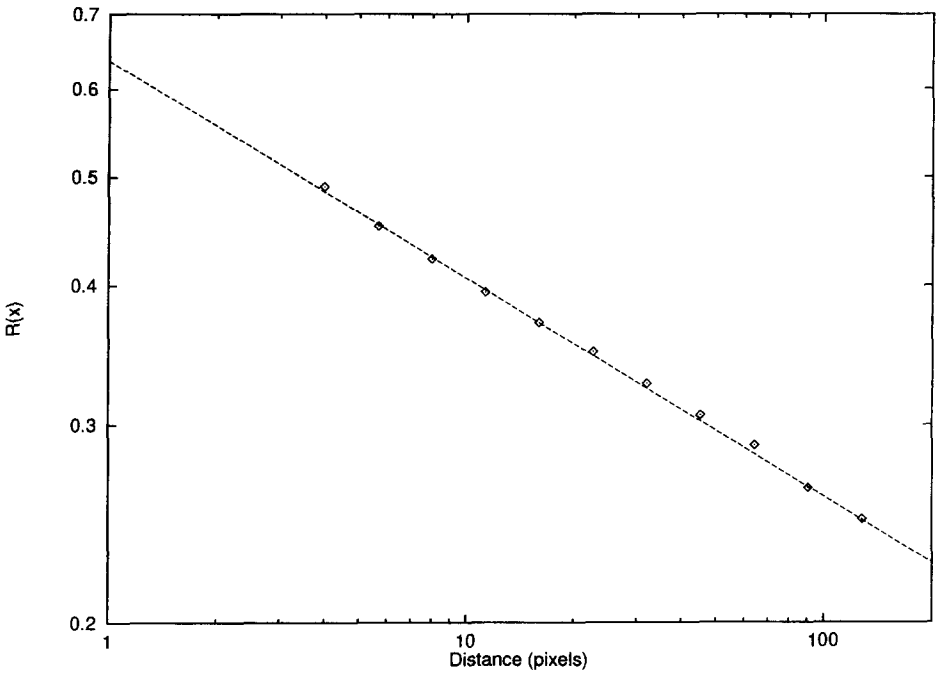

FIGURE 2. The residual function $R(x) = D_1 - D(x)$ for natural scenes; both axes are logarithmic, and the best fit linear regression gives a slope $\eta = 0.19$.

remarkably good fit to a power-law with the expected exponent.

Calibration independence

One of the most compelling features in the data on natural image scaling is invariance to choice of calibration. Indeed, if calibration were crucial it is doubtful that scaling would ever have been found. Furthermore, neural responses in the visual system are non-linear, and so the statistical effects of passing images through this "recalibration" are clearly important for understanding visual system design (Laughlin, 1981; van Hateren & van der

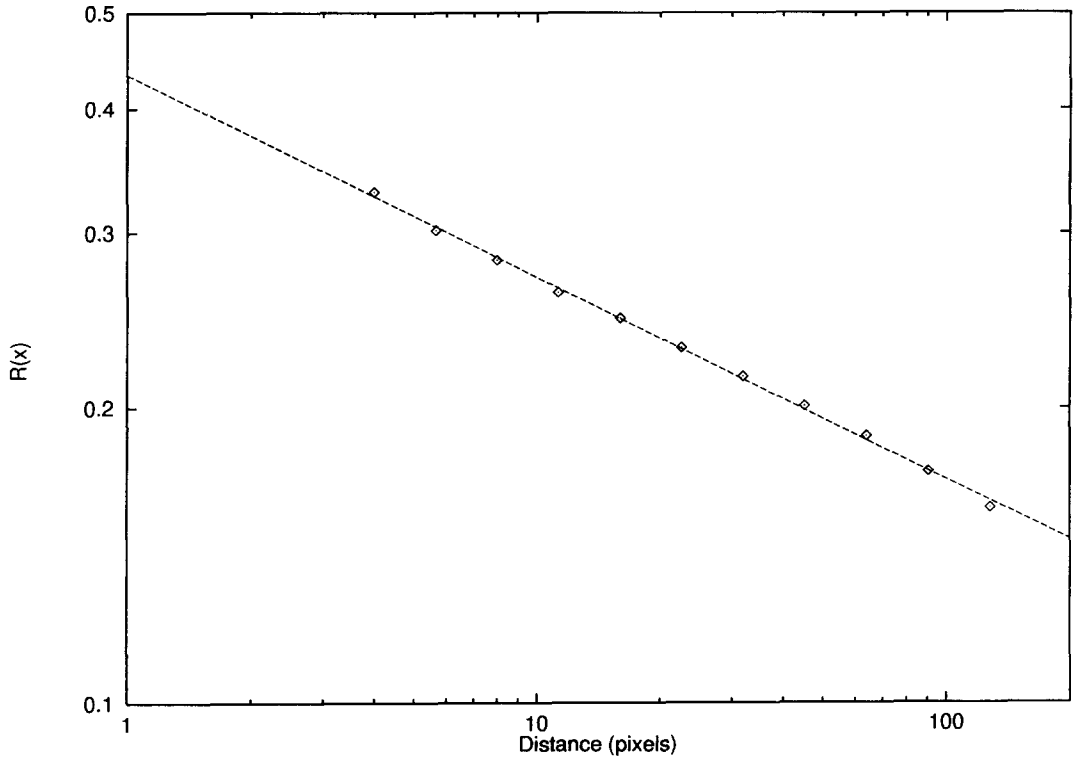

FIGURE 3. Residual function $R(x)$ for recalibrated images plotted on a log–log scale demonstrating its power-law nature. The best fit slope is $\eta' = 0.20$, only slightly different from the $\eta = 0.19$ value for the original data.

Schaaf, 1996). This section demonstrates that not only is calibration unimportant in different datasets, but that even recalibration of a single dataset fails to upset its own scaling.

As a simple experiment, a rather dramatic rescaling was preformed by converting all images in the dataset to black-and-white. Those pixels having a value greater than zero became white pixels, and the remainder became black. This recalibration produced a new dataset, on which the difference function $D(x)$ was recomputed. It again had the form of Eq. (6), though the best-fit exponent changed slightly from $\eta = 0.19$ to $\eta' = 0.20$. Given the drastic nature of the recalibration procedure, this change is surprisingly small. The residual function $R(x)$ for these images is plotted in Fig. 3.

This example is meant to demonstrate the robustness of scaling in natural images by pushing an extreme limit of recalibration. Of course it cannot be expected that entirely arbitrary recalibration (e.g., a random reassignment of pixel values) would preserve the correlational structure of the images. More reasonably, it probably holds as long as nearby pixel values generally remain nearby under recalibration. These ideas will be expanded upon shortly.

UNDERLYING REGULARITY: OBJECTS AND EDGES

What causes the pixel difference function $D(x)$ to take the form it does? Recall that it is found by choosing pairs of points within images and averaging the resulting squares of the pixel differences between those point pairs. I claim that each point pair can be considered as either belonging to the same object or to two different objects. This simple observation, though somewhat heuristic, can explain much about correlations in natural images.

Defining objects

The notion of objects is best grounded in the actual process of image generation, that is, the probability distribution of natural images. The model proposed in this study places randomly chosen objects in the world at random locations. The world is then illuminated and imaged, complete with occlusions. Since objects are chosen at random, pixels corresponding to different objects will have little statistical dependence on one another.* However, it is expected that pixels from the same object will be more closely related, if not only because of the likelihood of them originating from the same material and receiving similar lighting. How well this procedure actually mimics the environments of the datasets is unknown, but it will suffice as a first step. The model presented in the section entitled *"Static images"* will implement the procedure to yield exact results in this regime.

Suppose, then, that the difference function is being tabulated under such conditions. Point pairs separated by a distance x are chosen, and the pixel differences $\phi(0)-\phi(x)$ are squared and averaged together. There is a probability $P_{\text{same}}(x)$ that a given point pair of separation distance x belong to the same object. Recall that x is measured in degrees of visual angle, and so this probability depends on the actual spatial sizes of objects, their distribution of distances from the observer, and their shapes. Given that the two points belong to a single object there is then an associated difference function for same-object points separated by a distance x, $D_{\text{same}}(x)$ (isotropy is assumed). Similarly, if the points are from different objects then their difference function takes another form, say $D_{\text{diff}}(x)$.

By assigning all measurements to one of these two categories the overall difference function can be expressed as:

$$D(x) = P_{\text{same}}(x)D_{\text{same}}(x) + [1 - P_{\text{same}}(x)]D_{\text{diff}}(x). \quad (7)$$

The advantage of working in the spatial domain now becomes clear. This type of expression which comes from thinking about how images are composed does not have a simple analogy in Fourier space. Global frequency analysis does not provide good intuition about image structure whose character is fundamentally local.

But now the problem appears harder. Whereas before we had a single function $D(x)$, now we have three: $P_{\text{same}}(x)$, $D_{\text{same}}(x)$, and $D_{\text{diff}}(x)$. What we gain ultimately is insight into how images are made. Fortunately, as is demonstrated below, the problem actually becomes simpler when expressed this way. Suppose we wish to measure these functions. Instead of working with raw images we now have to "segment" them into objects and tabulate different sets of statistics, as well as the sizes of objects to get $P_{\text{same}}(x)$. The following section is an attempt to gather just such data.

Image segmentation: the difference functions

Each image in the database was segmented by eye into areas belonging to different objects. For example, the image shown in Fig. 1 was divided into regions corresponding to the stream, the rocks, the riverbank, the log on the riverbank, etc. Some less obvious decisions had to be made, such as identifying the whitecaps in the water as part of the stream as a whole, instead of defining them as independent. Leaves on trees were considered integral parts instead of objects in their own right (while their luminance may be significantly different from that of tree bark, they are statistically highly linked in appearance within the image). Suffice it to say that there is no entirely objective way of doing this. As it happened, the segmentation generally followed gross semantic boundaries. A more formal statistical definition will be presented in the section on image modeling.

Since objects can be differentiated down to microscopic scales, a cut-off of 5 pixels was chosen as a scale on which to define objects. A bed of leaves, each occupying 1 or 2 pixels in size, was considered a single object: "bed of leaves". If the leaves were each large enough to cover a 5×5 pixel square then they were considered independent objects. This procedure intro-

*Issues such as shadows which cover both objects, as well as the likelihood of nearby objects in the real world being related are no doubt important, but will be ignored in this elementary treatment.

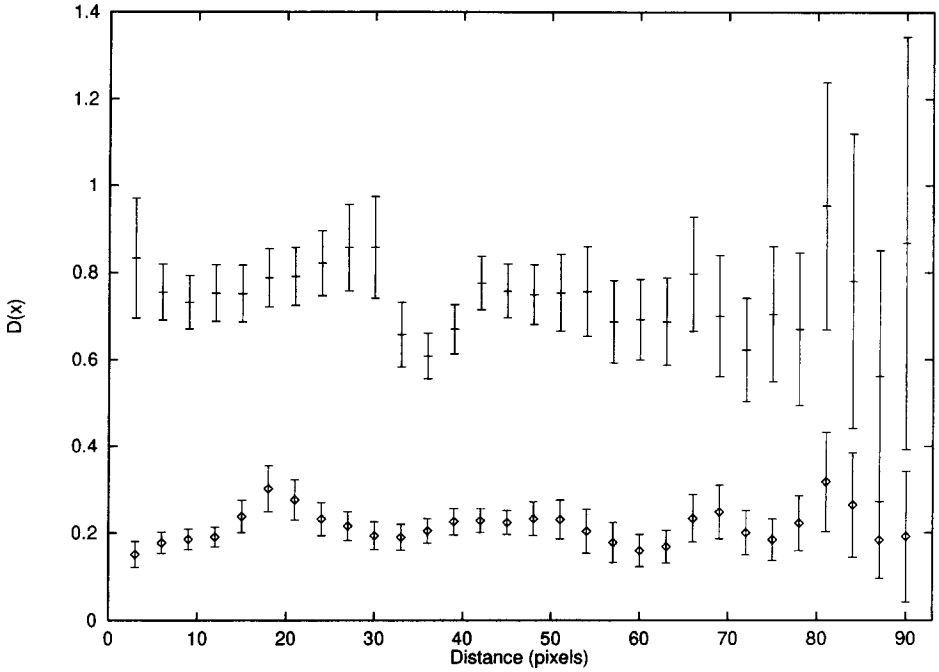

FIGURE 4. The difference functions $D_{\text{diff}}(x)$ (upper) and $D_{\text{same}}(x)$ (lower). Neither obeys a significant linear trend, and so they can be approximated by constants.

duced a "scale of uniformity" of 5 pixels, and so measurements below this scale have a separate meaning and were excluded. Some image regions were too dense with small branches and background to be segmented and were also excluded. In total, 1000 examples each of same and different object pixel regions were identified.* From each of these region pairs all possible pixel pairs were taken as data. Thus, two 5×5 image patches would produce 625 items of data (25×25). But since all of them originated in the same image regions, they do not represent independent data (of which we only really have 2000 examples). As such, during the estimation of standard errors the number of independent measurements (and not individual pixel pairs) was taken as the number of degrees of freedom.

Data are shown in Fig. 4. Two sets of datapoints are plotted. The upper ones correspond to $D_{\text{diff}}(x)$, and the lower ones are $D_{\text{same}}(x)$. The horizontal axis measures pixel separation. Note that the standard error bars are sizeable. On average, the number of independent region measurements contributing to each data point was nearly $n = 100$, so the data clearly involve large fluctuations. Next, all the data from $D_{\text{diff}}(x)$ are larger than those for $D_{\text{same}}(x)$. Simply put, pixels in different objects have a larger mean squared difference than those from the same object. This makes intuitive sense. More interestingly, within error bars neither of these graphs deviates significantly from a constant. Further, the separation between the graphs is somewhat larger than the fluctuations within graphs, thus providing another scale for comparison. Simply put, there are no obvious trends in the data.

Why might this be? Suppose two points belong to different objects. Then it does not really matter how far apart those two objects are separated, the pixels are always very much uncorrelated. A small effect might be similarity of illuminant for nearby objects, and so we expect a slight rise in the graph with distance. Similarly, for points within the same object, say a tree trunk, the separation distance is only marginally important. For instance, the bark on a tree remains at nearly the same luminance regardless of location on the trunk (again, modulo illuminant changes). So there is a reasonable expectation that the data appear the way they do.

Implications

The estimates for the constant difference functions are:

$$D_{\text{diff}} = 0.75 \pm 0.02$$

$$D_{\text{same}} = 0.22 \pm 0.01 \tag{8}$$

Rewriting Eq. (7) in terms of these constants gives:

$$D(x) = D_{\text{diff}} - (D_{\text{diff}} - D_{\text{same}})P_{\text{same}}(x), \tag{9}$$

while experimentally, $D(x)$ was found to take the form:

$$D(x) \approx 0.79 - 0.64x^{-0.19}. \tag{10}$$

Equating the measured and estimated difference function gives:

$$P_{\text{same}}(x) = a + bx^{-0.19}, \tag{11}$$

where a and b are constants. That is, the probability of remaining inside a given object at a separation distance x has a power-law component in the distance. Here we see

*The less biased alternative of allowing the computer to select these areas at random is not in practice workable. Too many of these random regions cannot be clearly identified as belonging to single objects. As such, I chose a somewhat biased, but reproducible, procedure.

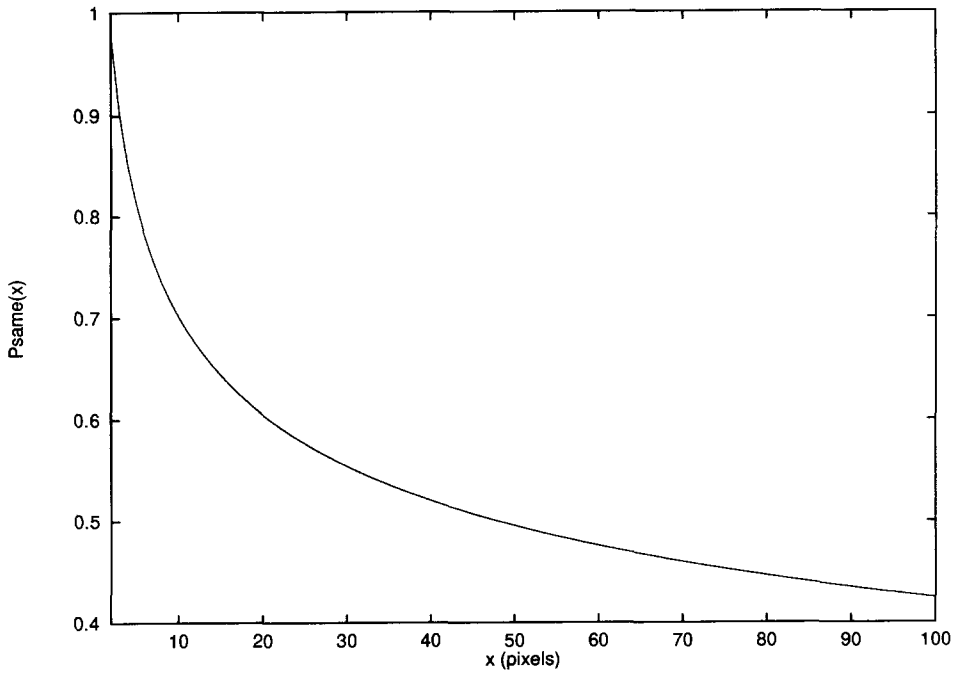

FIGURE 5. The form of $P_{\text{same}}(x)$ for the image ensemble as given by Eq. (12). Note the sizable probability of remaining in the same object even at separation distances of 100 pixels.

the origins of scaling in an occluding object model: it is the visible sizes of the objects themselves. The fact that many datasets show scaling implies that the manner in which objects occlude each other within many different environments universally produces a power-law such as Eq. (11).

From the measured values of D_{same} and D_{diff}, as well as the measured difference function of Eq. (10), the constants a and b can be found, giving a probability:

$$P_{same}(x) = -0.08 + 1.21x^{-0.19}, \qquad (12)$$

which is plotted in Fig. 5. This probability is invalid for short distances (less than 2 pixels or so) since it is larger than unity there. The objects previously defined were on scales greater than 5 pixels, and so it is not feasible to draw conclusions about objects on scales smaller than this. Had objects been discriminated on a smaller scale the constants within Eq. (12) would have been different. They reflect an intrinsic scale of measurement just as the multiplicative coefficient in the power spectrum reflects the spatial cut-off scales. At very large x this probability function also becomes invalid, as it becomes negative; a straightforward computation shows that this occurs at angular separations much larger than 360 deg, and it is thus unimportant.

The classical problem of image segmentation is that of identifying the borders of objects. Many techniques have been invoked to solve this very important problem in image processing (for a review see Pal & Pal, 1993). Some recent algorithms take into account local image statistics within and between objects in order to make segmentation decisions Bongiovanni *et al.*, 1993, 1994. A characterization of inter- and intra-object statistics is needed to optimize these approaches and to estimate the frequency of segmentation errors while using a given algorithm. Furthermore, these statistics are the starting point for the development of optimal algorithms.

Calibration independence (revisited)

Calibration independence of scaling follows naturally from Eq. (9). If we change the calibration of images, that is, the definition of ϕ(x) then the objects within each image do not change in location, shape, or identity. Furthermore, the relative constancy of the functions $D_{\text{same}}(x)$ and $D_{\text{diff}}(x)$ is not expected to alter either; they should remain flat. Their values will merely differ in a manner reflecting the recalibration. Thus, the only difference we expect in Eq. (9) is in the coefficients. Scaling will remain present, and the exponent will not change. This is exactly what was demonstrated at the end of the section entitled "*Calibration independence*".

To test this further, statistics were categorized at the same image locations identified earlier as either belonging to the same object or to different ones, but now on the recalibrated images mentioned earlier. This again provides two difference functions which are plotted in Fig. 6. Once again, the difference functions are essentially flat, with $D_{\text{diff}} > D_{\text{same}}$. Recalibration has not affected the basic form of Eq. (9).

Since recalibration does not affect the locations of objects within images, $P_{\text{same}}(x)$ is the same regardless of exactly how the images are calibrated or even the spectral sensitivity of the instrument making the measurements. Furthermore as the intra- and inter-object difference functions are constant, independent of calibration, we expect the scaling form of the difference function to be very robust, as it indeed appears to be.

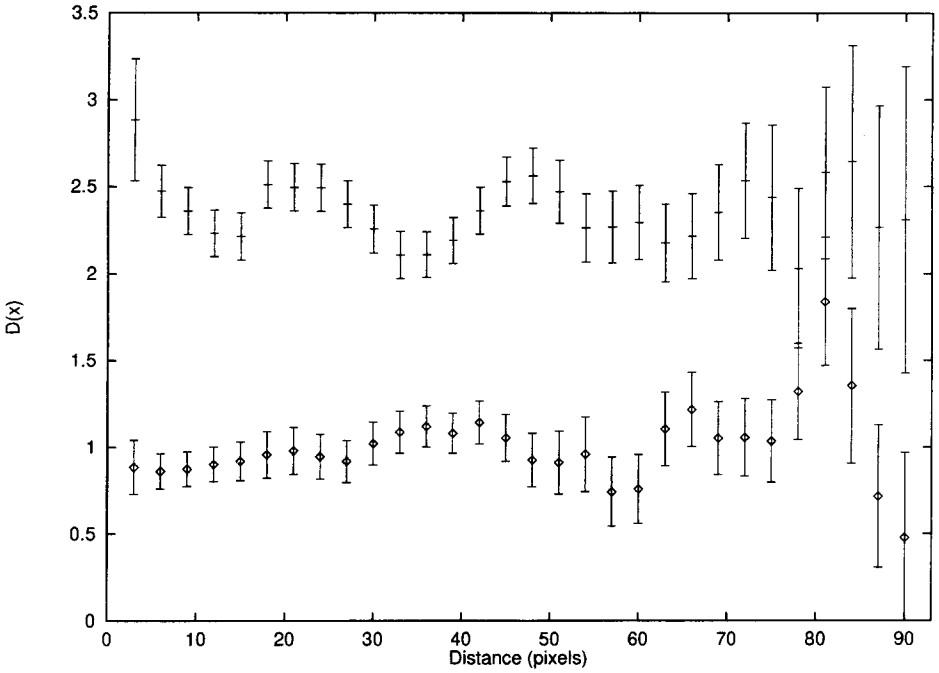

FIGURE 6. The difference functions $D_{\text{diff}}(x)$ (upper) and $D_{\text{same}}(x)$ (lower) for the same image locations as Fig. 4, but with the images recalibrated to black-and-white pixel values. The difference functions remain essentially flat to within error bars.

MODELING NATURAL IMAGES

The statements made so far constitute a plausible hypothesis. The claim is that the statistically independent components of images can be identified as "objects" which appear independently of one another in the natural environment. To obtain scaling in natural images those objects must have a viewable size distribution which obeys a power-law in angular separation. The data presented in the previous section were meant to supply evidence for this conclusion.

This section demonstrates that a simple planar model of occluding objects can be solved to provide this power-law distribution if the original objects are power-law distributed. That a power-law object distribution placed by hand produces power-law correlations within images is perhaps not surprising. However, this model allows us to generate scale-invariant images of given power spectrum in the *spatial* instead of the more usual Fourier domain, giving more control over the image structure. Furthermore, the model explicitly includes occlusion of one object by another, an important factor which is absent in those image models which use superposition such as linear independent components analysis (Olshausen & Field, 1996; Bell & Sejnowski, 1997). The model also explicitly demonstrates that scaling in images is not simply due to the $1/k^2$ spectrum of individual edges within images, thus laying to rest this popular notion.

Static images

Imagine walking on an infinite image plane. At a random location you blindly select from a number of choices an infinitesimally thin cardboard "cut-out" of some shape. You paint it a gray tone chosen from a distribution, and then drop it on the ground. This done, you continue to another random location and repeat the process.

This simple procedure creates images whose statistics can be easily assessed. The reason for studying such a model is that it explicitly identifies statistically independent objects and these objects occlude one another as they do in the real world. In these images the true "independent components" are the objects themselves, which have random size, location, and intensity. The model's most important deficiency is its two-dimensionality, whereas real images are projected and occluded in three dimensions. A treatment of this more complex and realistic case will have to await future study.

According to the previous section sufficient conditions for scaling of correlations within images are (a) that the probability distribution of not crossing an object border scale in distance; and (b) that objects have nearly uniform correlation within their borders and between different objects. Painting the objects in the model randomly assures condition (b), since correlations are uniform within the objects and zero between them. I show below that condition (a) is satisfied if the objects themselves are chosen with a power-law distribution of sizes. The form of this distribution explicitly determines the correlation function of the image ensemble, which can be chosen arbitrarily.

The correlation function of the model will be:

$$C(x) = C_0 P_{\text{same}}(x), \tag{13}$$

where C_0 is the constant correlation within objects, and the term for different objects is absent since they have zero correlation. Once again, a power-law in the correlation function implies a power-law in the spectrum.

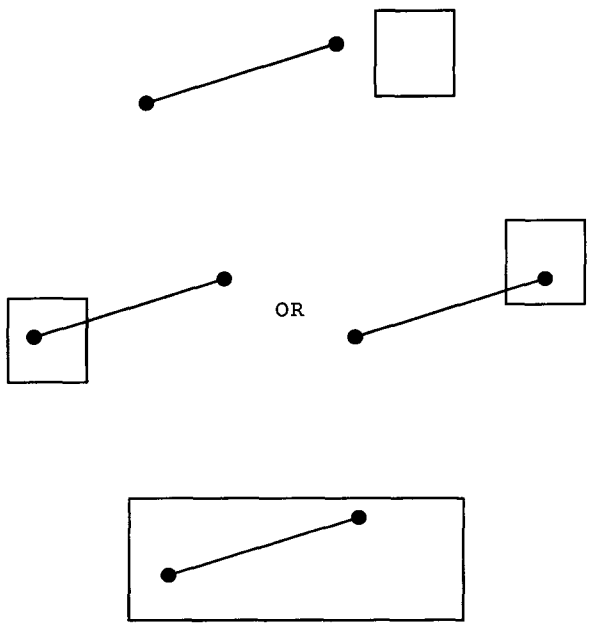

FIGURE 7. Possible relationships between rectangular objects and a pair of points separated by a distance x. Top: object contains neither point. Middle: object contains one point or the other. Bottom: object contains both points.

To obtain this scaling we must determine $P_{\text{same}}(x)$ itself to be power-law. The form of this function is derived below.

Imagine two fixed points in the plane a distance x apart. What is the probability that they are both within the same object in one of these randomly created images? Consider each of the flat objects in turn, starting with the most recently placed one. It may be considered to be "above" the rest, as it is not occluded by any others. There are three possibilities: (1) only one of the two points is within that object; (2) both points are situated within the object; and (3) neither point lies within the object (see Fig. 7). Let $p_1(x)$ be the probability that condition (1) occurs, $p_2(x)$ be the probability that condition (2) occurs, and $p_0(x)$ be the probability that condition (3) occurs (note that they sum to unity).

If either condition (1) or (2) occurs, then we know the answer for that particular image. If condition (3) arises then the object and the two points are entirely disjointed. We may simply ignore that object and repeat the procedure for the next one down the "stack" of objects. Thus, the probability of two points being in the same object may be determined recursively as:

$$P_{\text{same}}(x) = p_2(x) + p_0(x)P_{\text{same}}(x). \tag{14}$$

Here the probability is the sum of the probability that the highest object contains both points plus the probability that it contains neither point times the probability we seek.* Using $1-p_0 = p_1 + p_2$, we have:

$$P_{\text{same}}(x) = \frac{p_2(x)}{p_1(x) + p_2(x)}. \tag{15}$$

The problem is reduced to evaluating the functions $p_1(x)$ and $p_2(x)$.

For illustrative purposes, consider the simplest possible scaling images of this type, where all objects are circular and distributed in size according to the power-law:

$$P_s(s > s_0) = \frac{C}{s_0}\left(\frac{s}{s_0}\right)^{-\alpha}, \tag{16}$$

with the diameter $s > s_0$, some short-distance cut-off. With $\alpha > 1$ this distribution is integrable to infinite-sized objects. In fact, $\alpha > 3$ is required to make the model well-behaved, as the second moment of P_s must exist to evaluate the probabilities used below. By using a power-law distribution of object sizes we will obtain a power-law for $P_{\text{same}}(x)$.

It can be shown for these circles placed at random that:

$$p_1(x) = 2C(s_0/L)^2\left[\frac{1}{\alpha-3} - B(\alpha)(x/s_0)^{-(\alpha-3)}\right] \tag{17}$$

$$p_2(x) = C(s_0/L)^2 B(\alpha)(x/s_0)^{-(\alpha-3)},$$

where L is the image size. Here $B(\alpha) = \int_1^\infty du\, u^{2-\alpha} g(1/u)$ where $g(u)$ is a geometrical factor for the fraction of area within a unit-diameter circle which can be occupied by a point given that another point a fixed distance u away is also within the circle:

$$g(0 \le 1) = \frac{2}{\pi}\left[\cos^{-1}(u) - u\sqrt{1-u^2}\right]. \tag{18}$$

Using Eq. (15) gives:

$$P_{\text{same}}(x) = \frac{B(\alpha)(x/s_0)^{3-\alpha}}{2/(\alpha-3) - B(\alpha)(x/s_0)^{3-\alpha}}. \tag{19}$$

For $x \gg s_0$ the probability becomes a power-law:

$$P_{\text{same}}(x \gg s_0) = \frac{\alpha-3}{2}B(\alpha)\left(\frac{x}{s_0}\right)^{-(\alpha-3)}. \tag{20}$$

Thus, the scaling of inter-object probability follows directly from the scaling of apparent object sizes. In images of the real world this apparent size (in degrees) depends on an object's actual size as well as its distance from the observer. The overall distribution of apparent object size is thus a function of the distributions of object sizes and that of their distances. The preceding example is meant to demonstrate that the model can be solved and that it can give a power-law spectrum.

The above result was determined for pixel values associated with different objects being statistically independent. In fact, each object could display a texture on its surface without changing the essential result of Eq. (6), as long as the ensemble power spectrum of the

*Another way to arrive at this result is to imagine that we add another object to a given image. $P_{\text{same}}(x)$ must be invariant to this addition, since we still have an image of the same class. The probability of two points lying within the same object after the addition of another object is the probability that the new object contains both points ($p_2(x)$) plus the probability that the new object did not cover either point and both points were previously within the same object ($p_0(x)P_{\text{same}}(x)$). Thus, we have the same result as before.

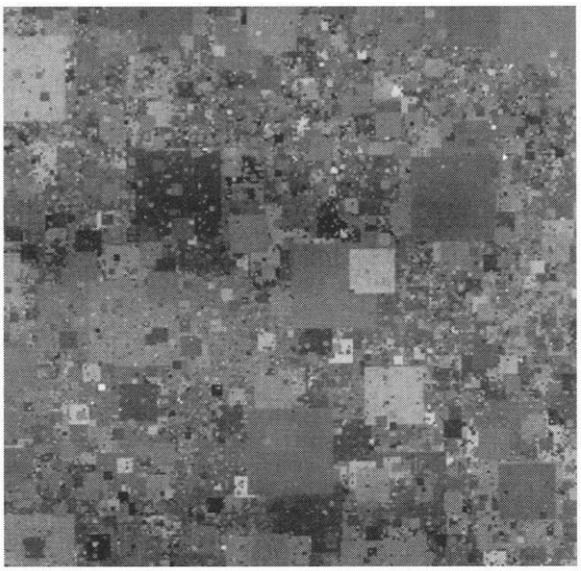

FIGURE 8. An image made of occluding squares whose size distribution is chosen to give the ensemble correlation function a power-law exponent $\eta = 0.19$.

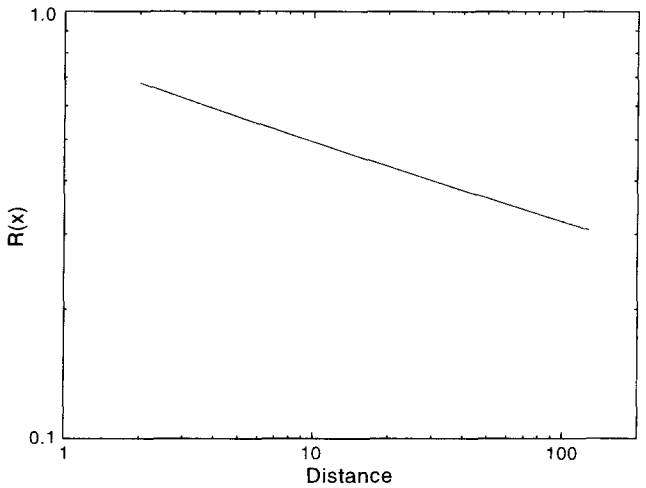

FIGURE 10. The orientationally-averaged correlation function (log–log scale) for the ensemble from which image 8 is drawn (slope = −0.19). This correlation function is derived in a similar manner to that of the section *Static images*.

textures is relatively flat so that $D_{\text{same}}(x)$ is essentially constant. Recalibration also would not affect the result, as long as some correlation remains within the objects. Note finally that by choosing an appropriate object size distribution P_s and object pixel histogram one can use this image generation method to produce an image ensemble both of given spectrum and of given pixel histogram. This is certainly not a trivial matter if one creates the images in the Fourier domain by enforcing a particular power spectrum. If the phases are chosen randomly then upon transforming back to the spatial domain the pixel histogram will be very nearly gaussian, since many independent numbers are combined at each pixel. Control over both the pixel histogram and the spectrum is thus not simple to achieve.

It has often been suggested that the origin of scaling in natural images is due to their being composed of luminance edges, each of which has a $1/k^2$ spectrum (see Carlson, 1978, for example). The occluding objects model does not support this conjecture. The important feature is not the characteristic form of object transitions (i.e., sharp edges), but rather the distribution of their occurrence as given by $P_{\text{same}}(x)$. This point is illustrated by presenting two images (Fig. 8 and Fig. 9), each of which consists entirely of edges, but are derived from ensembles with different two-point correlation functions. The image in Fig. 8 comes from an ensemble with power-law correlation, whereas the one in Fig. 9 has exponential correlations. These correlation functions are shown in Fig. 10 and Fig. 11, and can be derived from the same model as before. The choice of correlation function is made by selecting the functional form of P_s, the probability of object sizes, to be either power-law or exponential. The gray values of the squares were chosen independently from a Gaussian distribution.

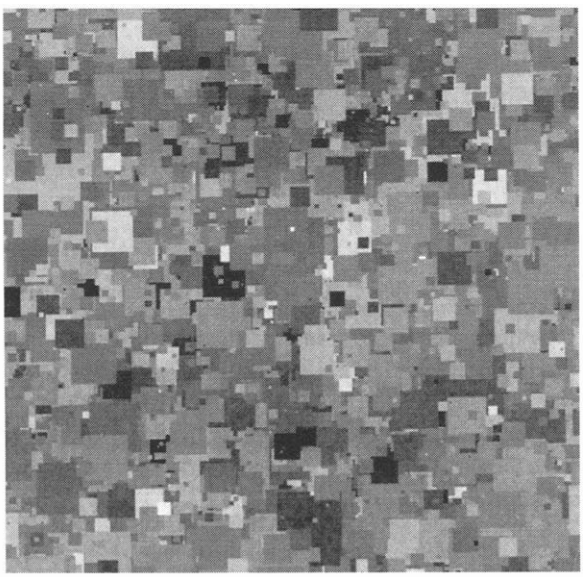

FIGURE 9. An image made of occluding squares whose size distribution is chosen to give the ensemble correlation function an exponential form.

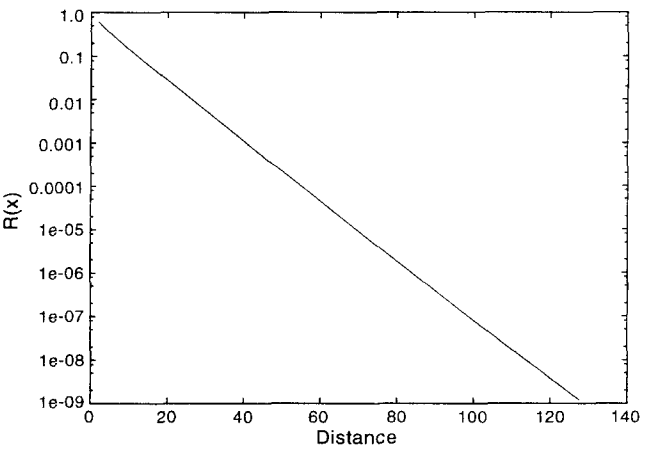

FIGURE 11. The orientationally-averaged correlation function (semi-log scale) for the ensemble from which image 9 is drawn (1/e height at 6.2 pixels).

Moving images

Recent measurements on natural "movies" by Dong & Atick (1995a) have demonstrated that their correlations take the particular scaling form:

$$S(k,\omega) \propto \frac{1}{k^{3-\eta}} F(\omega/k), \tag{21}$$

where S is the power spectrum, k is the magnitude of the spatial frequency, ω is the temporal frequency, and $2-\eta$ is the spectral exponent for static scenes. The linkage between spatial and temporal frequencies demonstrated in the functional form $F(\omega/k)$ implies a scaling relation between the two. Dong and Atick argue that since the unit of their quotient is a velocity the dynamics of these scenes can be understood physically in terms of object motion. They proceed to derive this scaling form by considering images of a given static spatial spectrum moving at a random velocity chosen from a distribution. The resulting spatio-temporal spectrum contains a function $F(\omega/k)$ which is specified by the choice of velocity distribution.

The simple assumption of a scaling distribution implies that the spectrum must be of the form:

$$S(k,\omega) = k^{-\beta} Q(\omega k^{-\beta}) S_0(k), \tag{22}$$

where S_0 is the static spectrum of Eq. (1), and $Q(u)$ is a distribution which integrates to unity ($\int \frac{du}{2\pi} Q(u) = 1$). That this integration should yield the static spectrum follows from the fact that taking a snapshot adds together the power in all of the temporal frequency band. The exponent β depends on the exact nature of the dynamic scaling. The fact that Dong and Atick find $\beta = 1$ means that object velocity is a fundamental quantity.

Can the method of dividing the visual world into uncorrelated objects reproduce the result? First, the Dong and Atick result is expressed in terms of spatial correlations (through a calculation similar to that in the Appendix) as:

$$C(x,t) = A + Bx^{-\eta} G(x/t), \tag{23}$$

again assuming spatial isotropy. Extending the original model to include the time domain, circular objects are considered, whose sizes are chosen from the distribution of Eq. (16), whose velocities are chosen randomly from an isotropic distribution $Pv(v)$, and whose gray values are uncorrelated between objects. Just as before each object lay in its own infinitesimal plane, so now do they each move with constant velocity within these planes. This model's correlation function is given by:

$$C(x,t) = C_{\text{same}} P_{\text{same}}(x,t), \tag{24}$$

where $P_{\text{same}}(x, t)$ is the probability that two points separated by a distance x in space and a delay t in time belong to the same object.

As before we must seek the two functions $p_1(x, t)$ and $p_2(x, t)$, the probabilities that a randomly placed moving object contains only one of two points separated by (x, t), or both points, respectively. The relation of Eq. (15) is again used to determine $P_{\text{same}}(x, t)$. Since the size of an object and its velocity are chosen independently, the spatio-temporal probabilities can be expressed as:

$$p_1(x, t) = \langle p_1(|\mathbf{x} - \mathbf{v}t|) \rangle_{\mathbf{v}}, \tag{25}$$

and similarly for p_2.

We start first with the expression for p_2, which requires an integration over the magnitude of the velocity and its orientation:

$$p_2(x,t) =$$

$$aCs_0^2 B(\alpha) \int_0^\infty dv\, v P_v(v) \int_0^{2\pi} d\theta (|\mathbf{x} - \mathbf{v}t|/s_0)^{3-\alpha}$$

$$= aCs_0^2 B(\alpha)(x/s_0)^3 - G(x/t), \tag{26}$$

where $G(v)$ is a function which depends on the velocity distribution $Pv(v)$, and $B(\alpha)$ is as before. For p_1 we have:

$$p_1(x,t) = 2aCs_0^2 \left[\frac{1}{\alpha - 3} - B(\alpha)(x/s_0)^{3-\alpha} G(x/t) \right]. \tag{27}$$

Combining these functions gives:

$$P_{\text{same}}(x,t) = \frac{B(\alpha)(x/s_0)^{3-\alpha} G(x/t)}{2/(\alpha - 3) - B(\alpha)(x/s_0)^{3-\alpha} G(x/t)}. \tag{28}$$

In the limit $x \gg s_0$ we arrive at:

$$P_{\text{same}}(x,t) = \frac{\alpha - 3}{2} B(\alpha) \left(\frac{x}{s_0} \right)^{-(\alpha-3)} G(x/t). \tag{29}$$

This is precisely the form required to satisfy Eq. (23).

Once again, the objects could be textured or be of different shapes without affecting the result. A primary restriction, however, is the motion of objects only within their own planes. In the real world the looming of objects due to motion of the observer, as well as changes in object velocity are both highly salient. However, reproducing the second-order statistics of real images can be achieved with this simple model. It is the higher-order ones which describe more precisely the structure of natural images.

CONCLUSIONS

The first goal of this paper has been to understand why so many disparate measurements all draw the same conclusion: natural images scale. The cause of this scaling must be robust, both to changes in environment and to recalibration. The answer proposed here is that natural images are composed primarily of statistically independent objects which occlude one another. Further, natural environments tend to arrange themselves so that the image regions corresponding to these objects are power-law in size. Combined, these two properties give rise to scaling universally. It should be noted that other models of natural images (e.g., based on superposition; Field, 1994) can also account for scaling. The advantage of the current model is its ability to capture the process of

object occlusion and to explain the calibration independence of scaling.

Scaling is usually expressed in terms of the power spectrum. However, as shown in the Appendix, the result is equally well expressed in the spatial domain in terms of the correlation function or the difference function defined here. What this reformulation provides best is some intuition toward the origins of scaling. It has been proposed in this study that the measurement of the difference function corresponds to two disjointed sets: data from same object point pairs, and data from different object pairs. Scaling then follows from the probability distribution of object crossings as a function of separation distance. The final explanation for scaling in natural images must then rest on understanding why many different image environments display scaling in apparent object sizes. It appears that these environments consistently arrange themselves to obey power-law statistics, much like those found in self-organized critical systems (Bak & Paczuski, 1995).

A simple measurement performed on a natural image database showed that the difference functions within and between objects are both relatively flat. This distance independence does not falter under recalibration, thus lending support to the "images from objects" model of natural scenes. Finally, a two-dimensional image model was demonstrated which can be selected to give any ensemble power spectrum (by choosing the distribution of object sizes) and any image histogram. This degree of control is not readily available by synthesizing images purely in the Fourier domain. The model was used to show that edges are not the essential features within images which cause scaling.

The model presented in this study is too simple to model imaging of the real world completely accurately. To do so requires a truly three-dimensional model of moving objects (or perhaps of a changing point-of-view) bathed in an illuminant which scatters off surfaces. A basic analytical result still awaits, though progress could surely also be made simply through rendering naturalistic environments by computer. The robustness of scaling could then be tested in a number of model environments created using various parameters, such as object size distributions and texture maps. The near photo-reality of contemporary computer graphics could well save much experimental effort in this endeavor.

It should be noted that the only statistics I have attempted to model are those of second-order. At the same time it is also known that higher-order statistics also show scaling, such as those of pixel difference histograms (Ruderman & Bialek, 1994b; Ruderman, 1994b). This deeper scaling places further restrictions on image models used to explain natural image statistics, and itself remains to be illuminated theoretically. The notion of statistically independent image patches belonging to different objects does, however, imply that there is a single scaling exponent in these images. An explanation for why the power spectrum of local variance images has the same exponent as the spectrum of the images themselves (Ruderman, 1994b) thus appears to be at hand.

Finally, the understanding we gain about the images our visual system interprets ultimately helps us to understand the process of vision itself. Barlow's notion of optimally in non-redundant representations and factorial coding is, in practice, linked to the statistically independent objects present in the world, and their optical projections as visual stimuli. Combining efficient coding with the visual ecology of image statistics offers the promise of a unified approach to the systematic exploration of visual system design.

DERIVING THE CORRELATION FUNCTION

In this Appendix the form of the correlation function from the power-law spectrum is derived in detail. Since the results of the calculation are important and not obvious at first sight this calculation is included for completeness. Those with little mathematical inclination may wish simply to view the table of results at the end.

The theory of wide-sense stationary processes (Papoulis, 1991) states that the power spectrum and the correlation function are related through the Fourier transform:

$$S(k) \Leftrightarrow C(x). \tag{A1}$$

In two dimensions we have the precise relations:

$$S(\mathbf{k}) = \int d^2x C(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}$$

$$C(\mathbf{x}) = \int \frac{d^2k}{(2\pi)^2} S(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{A2}$$

where the integrals are over all space and frequencies, respectively. The frequency variable, k, is measured in 2π cycles/deg, and the spatial variable, x, is measured in degrees. For the purposes of this Appendix, we will be interested in the second equation of this pair.

The power spectrum of natural images takes the form $S(k) = Ak^{\eta-2}$, where any orientational dependence has been ignored. The corresponding correlation function will also be of an orientationally averaged nature. The orientation dependence of Eq. (A2) can then be integrated out, yielding a Hankel transform:

$$C(x) = \frac{A}{2\pi} \int dk k^{\eta-1} J_0(kx). \tag{A3}$$

The limits of this k integral must be examined. Using the following small and large argument limits of the Bessel function [Gradshteyn & Ryzhik, 1994; (Eqs 8.441.1) and (8.451.1)]:

$$J_0(u \longrightarrow 0) \longrightarrow 1$$

$$J_0(u \longrightarrow \infty) \longrightarrow \sqrt{\frac{2}{\pi u}} \cos(u - \pi/4), \tag{A4}$$

we see that (for finite x) the $k \to 0$ limit of the integral is non-divergent for $\eta > 0$. The upper limit is well-behaved for $\eta < 3/2$, as will be discussed shortly. So for $0 < \eta < 3/2$ the integral converges and no cut-offs need be imposed. Of course the process of imaging itself always introduces a high-frequency roll-off, and since arbitrarily low spatial frequencies do not exist (what does one cycle per 720 deg mean when the image wraps around at 360 deg?), we will generally need to introduce cut-offs. Since no η has yet been found as large as 3/2 this limitation is not expected to be significant. The high-frequency cut-off will turn out to be irrelevant in this regime, as long as it is much larger than $1/x$. As a first-pass result, taking the complete range of frequencies as opposed to imposing cut-offs gives for Eq. (A3):

$$C(x) \propto x^{-\eta} \tag{A5}$$

(valid for $0 < \eta < 3/2$), thus showing immediately that we expect the correlation function to include a power-law along with cut-off-dependent terms.

The preceding paragraph was meant to motivate the need to introduce high- and low-frequency cut-offs, which we define as $\lambda < k < \Lambda$. Solving for the correlation function begins in earnest by expressing it in two pieces:

$$C(x) = \frac{A}{2\pi} x^{-\eta}[I_\lambda(x) - I_\Lambda(x)], \tag{A6}$$

with

$$I_\lambda(x) = \frac{1}{2\pi}\int d\theta \int_{\lambda x}^{\infty} du\, u^{\eta-1} e^{iu\cos\theta} \tag{A7}$$

being the integral from the low cut-off to infinite frequency, and similarly for $I\Lambda(x)$. I will assume for the remainder of the discussion that $\lambda x < \Lambda x$; that is, the length scales over which we wish to measure the correlation function are far from those corresponding to the cut-off frequencies. Let us examine each of these integrals in turn.

Using an expression for the incomplete gamma function [Gradshteyn & Ryzhik, 1994; Eq. (3.381.3)] gives:

$$I_\lambda(x) = \frac{1}{2\pi}\int d\theta(-i\cos\theta)^{-\eta}\Gamma(\eta, -i\lambda x\cos\theta). \tag{A8}$$

As $\lambda x \to 0$ the incomplete Gamma function can be expanded to lowest order as $\Gamma(\eta, q\to 0) \to \Gamma(\eta) - q^\eta/\eta$ [Gradshteyn & Ryzhik, 1994; Eq. (8.354.2)], giving:

$$I_\lambda(x) = -\frac{(\lambda x)^\eta}{\eta} + \frac{\Gamma(\eta)}{2\pi} i^\eta \int d\theta(\cos\theta)^{-\eta}. \tag{A9}$$

The θ integral can now be done [Gradshteyn & Ryzhik, 1994; Eq. (3.621.1)]:

$$\int_0^{2\pi} (\cos\theta)^{-\eta} = 2^{1-\eta} e^{\pi i\eta/2} \cos(\pi\eta/2) \frac{\Gamma^2\left(\frac{1-\eta}{2}\right)}{\Gamma(1-\eta)}. \tag{A10}$$

This gives our final expression (valid for $\eta \ll 3/2, \lambda x \ll 1$):

$$I_\lambda(x) = \frac{2^\eta}{\pi}\cos(\pi\eta 2)\frac{\Gamma(\eta)\Gamma^2\left(\frac{1-\eta}{2}\right)}{\Gamma(1-\eta)} - \frac{(\lambda x)^\eta}{\eta}. \tag{A11}$$

Using the large distance behavior of the Bessel function as given by Eq. (A3), we can write I_Λ in the limit $\Lambda x \to \infty$ as:

$$I_\Lambda(x) = \frac{1}{\sqrt{\pi}}\int_{\Lambda x}^{\infty} du\, u^{\eta-\frac{3}{2}}[\cos u + \sin u]. \tag{A12}$$

The cosine and sine integrals can be performed [Gradshteyn & Ryzhik, 1994; Eq. (3.761)] to give:

$$I_\lambda(x) = \frac{1}{\sqrt{2\pi}}\left[e^{-\pi i(\eta-\frac{1}{2})/2}\Gamma(\eta - \frac{1}{2}, i\Lambda x) + c.c.\right], \tag{A13}$$

where "c.c." denotes the complex conjugate of the same expression. Expanding the incomplete Gamma function in large Λx and keeping the dominant term [Gradshteyn & Ryzhik, 1994; Eq. (8.857)]:

$$\Gamma(\eta - \frac{1}{2}, q \to \infty) \to q^{\eta-\frac{1}{2}} e^{-q} \tag{A14}$$

gives:

$$I_\Lambda(x) = \sqrt{\frac{2}{\pi}}(\Lambda x)^{\eta-\frac{3}{2}}\sin(\Lambda x). \tag{A15}$$

For $\eta < 3/2$, as $\Lambda x \to \infty$ at fixed x, $I_\Lambda(x)$ becomes negligible whereas $I_\lambda(x)$ remains finite. Thus the high-frequency cut-off can be ignored for $x >> 1/\Lambda$, or when we measure the correlation function at distances greater than the pixel spacing. The final expression for the correlation function becomes (valid for $\lambda x \ll 1 \ll \Lambda x$, $\eta < 3/2$):

$$C(x) = \left(\frac{A}{2\pi}\right)\left[\frac{2^\eta}{\pi}\cos(\pi\eta/2)\frac{\Gamma(\eta)\Gamma^2\left(\frac{1-\eta}{2}\right)}{\Gamma(1-\eta)} x^{-\eta} - \frac{\lambda^\eta}{\eta}\right]. \tag{A16}$$

The case $\eta = 0$ is special, and the correlation function can be found from Eq. (A16) by a limiting procedure as:

TABLE 1. Behavior of the correlation function at various, η.

Regime	$C(x)$
$\eta > 0$	$-\lvert C_1\rvert + \lvert C_2\rvert x^{-\lvert\eta\rvert}$
$\eta = 0$	$-\frac{A}{2\pi}\ln(\lambda x)$
$\eta < 0$	$\lvert C_1\rvert - \lvert C_2\rvert x^{\lvert\eta\rvert}$

$$C(x) = -\frac{A}{2\pi}\ln(\lambda x). \tag{A17}$$

The correlation function for an exponent of exactly two is thus logarithmic.

To summarize, except for the case $\eta = 0$ which gives a logarithmic correlation function, the correlation function takes the form:

$$C(x) = -C_1 + C_2 x^{-\eta}, \tag{A18}$$

with C_1 and C_2 determined by Eq. (A16). The smoothness of the transition through the logarithm at $\eta = 0$ is well-controlled by the dependences of C_1 and C_2 on η. It can be shown that for η positive (negative) C_1 and C_2 are both positive (negative), thus always providing a correlation function which falls with separation distance. This is summarized in Table 1. Note that the value of $C1$ depends on the lower cut-off frequency λ.

REFERENCES

Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing? *Network, 3*, 213–251.

Atick, J. J., Li, Z. & Redlich, A. N. (1992). Understanding retinal color coding from first principles. *Neural Computation, 4*, 559–572.

Atick, J. J. & Redlich, N. (1990). Towards a theory of early visual processing. *Neural Computation, 2*, 308.

Atick, J. J. & Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation, 4*, 196–210.

Baddeley, R. J. (1994). Visual statistics using neural networks. PhD thesis, University of Stirling, Stirling, Scotland.

Bak, P. & Paczuski, M. (1995). Complexity, contingency, and criticality. *Proceedings of the National Academy of Sciences USA, 92*, 6689–6696.

Bell, A. J. and Sejnowski, T. J. (1997). The independent components of natural images are edge filters. *Vision Research, 37*, 3327–3338.

Bialek, W., Ruderman, D. L. and Zee, A. (1991). Optimal sampling of natural images: a design principle for the visual system? In Touretzky, D. & Moody, J. (Eds), *Advances in neural information processing systems 3* (pp. 363–369). San Mateo, CA: Morgan Kaufmann.

Bichsel, M. (1994). Segmenting simply connected moving objects in a static scene. *IEEE Transactions of Pattern Analysis & Machine Intelligence, 16*, 1138–1142.

Bongiovanni, G., Cinque, L., Levialdi, S. & Rosenfeld, A. (1993). Image segmentation by a multiresolution approach. *Pattern Recognition, 26*, 1845–1854.

Braddick, O., Campbell, F. W. and Atkinson, J. (1978). Spatial frequency channels in vision. In Held, R., Leibowitz, H. W. & Teuber, H-L. (Eds), *Handbook of sensory physiology VIII, Perception*. Heidelberg: Springer.

Burton, G. J. & Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics, 26*, 157–170.

Carlson, C. R. (1978). Thresholds for perceived image sharpness. *Photographic science and engineering, 22*, 69–71.

Coleman, P. H. & Pietronero, L. (1992). The fractal structure of the universe. *Physics Reports, 213*, 311–389.

Dong, D. W. & Atick, J. J. (1995a). Statistics of natural time-varying images. *Network, 6*, 345–358.

Dong, D. W. & Atick, J. J. (1995b). Temporal decorrelation—a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network, 6*, 159–178.

Field, D. J. (1987). Relations between the statistics of natural images

and the response properties of cortical cells. *Journal of the Optical Society of America A, 4*, 2379.

Field, D. J. (1993). Scale-invariance and self-similar wavelet transforms: an analysis of natural scenes and mammalian visual systems. In Farge, M., Hunt, J. C. R. & Vassilicos, J. C. (Eds), *Wavelets, fractals, and Fourier transforms* (pp. 151–193). Oxford: Clarendon Press.

Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation, 6*, 559–601.

Gradshteyn, I. S. and Ryzhik, I. M. (1994). *Table of integrals, series, and products*. New York: Academic Press.

van Hateren, J. H. (1992a). Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *Journal of Comparative Physiology A, 171*, 157–170.

van Hateren, J. H. (1992b). A theory of maximizing sensory information. *Biological Cybernetics, 68*, 23–29.

van Hateren, J. H. & van der Schaaf, A. (1996). Temporal properties of natural scenes. In Rogowitz, B. E. & Allebach, J. P. (Eds), *IS&T/SPIE Proceedings-Human vision and electronic imaging*, 2657, SPIE, pp. 139–143.

von Helmholtz, H. (1881). On the relation of optics to painting. In *Popular scientific lectures*. New York: Appleton.

Laughlin, S. B. (1981). A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsh, 36c*, 910–912.

Laughlin, S. B. (1992). Retinal information capacity and the function of the pupil. *Opthalmic and Physiological Optics, 12*, 161–164.

Li, Z. & Atick, J. J. (1994). Toward a theory of the striate cortex. *Neural Computation, 6*, 127–146.

Linsker, R. (1994). Sensory processing and information theory. In Grassberger, P. and Naadal, J.-P. (Eds), *From statistical physics to statistical inference and back* (pp. 237–247). Dordrecht: Kluwer.

Lythgoe, J. N. (1979). *The ecology of vision*. New York: Oxford University Press.

Olshausen, B. A. & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381*, 607–609.

Pal, N. R. & Pal, S. (1993). A review on image segmentation techniques. *Pattern Recognition, 26*, 1277–1294.

Papoulis, A. (1991). *Probability, random variables, and stochastic processes*, 3rd edn. New York: McGraw Hill.

Procaccia, I. (1984). Fractal structures in turbulence. *J. Stat. Phys., 36*, 649–664.

Ruderman, D. L. (1989). Undergraduate thesis. Univ. of California, Berkeley: Physics Department.

Ruderman, D. L. (1994a) Designing receptive fields for highest fidelity. *Network, 5*, 147–155.

Ruderman, D. L. (1994b) The statistics of natural images. *Network, 5*, 517–548.

Ruderman, D. L. (1996). Origins of scaling in natural images. In Rogowitz, B. E. & Allebach, J. P. (Eds), *IS&T/SPIE Proceedings-Human vision and electronic imaging*, 2657, SPIE, 120-131.

Ruderman, D. L. & Bialek, W. (1994). Statistics of natural images: scaling in the woods. In Cowan, J. D., Tesauro, G. and Alspector, J. (Eds), *Advances in neural information processing systems 6*. San Francisco, CA: Morgan Kaufmann.

Ruderman, D. L. & Bialek, W. (1994b). Statistics of natural images: scaling in the woods. *Phys. Rev. Lett., 73*, 814–817.

van der Schaaf, A. & van Hateren, J. H. (1996). Modelling the power spectra of natural images: statistics and information. *Vision Research, 28*, 2759–2770.

Srinivasan, M. V., Laughlin, S. B. & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London B, 216*, 427–459.

Tolhurst, D. J., Tadmor, Y. & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic and Physiological Optics, 12*, 229–232.

Turcotte, D. L. (1995). Scaling in geology: landforms and earthquakes. *Proceedings of the National Academy of Science USA, 92*, 6697–6704.

Vicsek, T. (1992). *Fractal growth phenomena*. Singapore: World Scientific.

Zocchi, G., Moses, E. & Libchaber, A. (1990). Coherent structures in turbulent convection, an experimental study. *Physica A, 166*, 387–407.

Acknowledgements—I am indebted to Horace Barlow and Bill Bialek for conversations which were instrumental to formulating the basic ideas presented in this work. My gratitude is extended to Giuliano Gavazzi for guidance in deriving Eq. (15) and to those who commented on the manuscript: Horace Barlow, Bill Bialek, Mike DeWeese and Bruno Olshausen. Special thanks are extended to Simon Laughlin for conversations in the Clare College gardens which revealed many occluded items. Work was supported by NSF NATO grant RCD-9353730.